

Title: Flax Seed Specific Promoters

This application is a continuation-in-part application of United States Patent Application No. 09/645,593 filed August 25, 2000 (now allowed), which claims benefit from United States provisional application no. 60/151,044, filed August 27, 1999 (now abandoned) and United States provisional application no. 60/161,722 filed on October 27, 1999 (now abandoned), all of which are incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates to plant genetic engineering methods useful for the alteration of the constituents of plant seeds. More specifically, the invention relates to promoters that have been obtained from flax and are capable of directing expression of non-native genes in flax seeds as well as the seeds of other plants.

BACKGROUND OF THE INVENTION

Flax or linseed (*Linum usitatissimum*) is a commercially important oilseed crop. Flax oil and meal are valuable raw materials derived from flax seed. A further economically significant raw material, flax fiber, is obtainable from the stem of the plant. The flax oil fraction is used for non-edible purposes, for example in the manufacture of varnish and paint, and has more recently become suited for use in the manufacture of a range of edible products, such as margarines and salad oils and dressings, by virtue of newly bred so called Linola cultivars (Green (1986) Can. J. Plant Sci, 66: 499-503). Flax meal is used primarily as a constituent of ruminant feeds while flax fibers are used in the manufacture of linen fabrics. Given its economic importance as a source for raw materials, it is desirable to further improve and diversify the available flax cultivar portfolio both with respect to agronomic performance, for example seed yield, resistance to pathogens and low climatic temperatures, and with respect to yield and quality of the raw materials to suit downstream applications. Although it is possible to obtain improved flax cultivars through conventional plant breeding, as evidenced by the development of the Linola cultivars, developing an elite agronomic plant line requires large

investments in plant breeding due to the long timelines involved. Plant genetic engineering technology allows the isolation of genes directly from unrelated species and the transfer of these genes into elite agronomic backgrounds, thereby significantly reducing the time required to develop new cultivars. In addition plant genetic engineering permits the manufacture of products not naturally obtainable from flax, for example therapeutic agents.

In order to develop novel flax cultivars through plant genetic engineering, control over the expression of the introduced foreign or non-native gene is of critical importance. The desired expression characteristics for the non-native gene, such as the level of expression of the non-native gene, the particular plant tissue or organ in which the non-native gene is expressed, and the particular time in the growth cycle of the plant at which the non-native gene is expressed, will vary depending on the application for which the plant line is developed. For example, the modification of the seed oil composition may require low levels of seed-specific expression of an enzyme involved in fatty acid metabolism at an early stage in seed development (see for example US Patent 5,420,034). On the other hand expression of a pharmaceutical protein could preferably require high levels of leaf-specific expression upon harvesting of the plant leaves (see for example, US Patent 5,929,304).

In order to manipulate the expression characteristics of non-native genes numerous factors can be influenced. One factor is the choice of the transcriptional promoter used. A wide range of plant compatible promoters is currently available and some of the better documented promoters include constitutive promoters such as the 35-S CaMV promoter (Rothstein *et al.* (1987), Gene 53: 153-161) and the ubiquitin promoter (US Patent 5,614,399), tissue specific promoters such as seed-specific promoters, for example the phaseolin promoter (Sengupta-Gopalan *et al.*, (1985), PNAS USA 82: 3320-3324) and inducible promoters, such as those inducible by heat (Czarnencka *et al.*, (1989), Mol. Cell. Biol. 9 (8): 3457-3464), UV light, elicitors and wounding (Lois *et al.*, (1989) EMBO J. 8 (6): 1641-1648), or chemicals such as endogenous hormones (Skriver *et al.* (1991), Proc. Natl. Acad.

Sci. USA 88(16): 7266-7270). Other factors that can be manipulated in order to control the expression characteristics of non-native gene in plants include transcriptional modification factors such as introns, polyadenylation sites and transcription termination sites. The expression characteristics of the non-native gene can further be manipulated by factors that affect translation, such as ribosomal binding sites and the codon bias that is exhibited by the host. Furthermore, the non-native gene itself may affect the viability of the transgenic plant, thus limiting particularly the levels of expression that can be attained. In some cases it may be possible to overcome this problem, by expressing the protein in a tissue specific manner, e.g. in the leaves or seed, or by restricting the accumulation of the protein in different subcellular compartments such as for example the cytoplasm, the endoplasmic reticulum or vacuoles, typically by the presence or the absence of specific targeting sequences capable of directing the protein to these compartments. Another factor that will affect the expression characteristics is the location in which the construct inserts itself into the host chromosome. This effect could provide an explanation as to why different plants, transformed with the same recombinant construct, can have fluctuating levels of recombinant protein expression.

To the best of the inventors' knowledge, expression of non-native genes in flax seeds is only documented in PCT Patent Application WO 98/18948. This application discloses two stearyl-acyl carrier protein desaturase (SAD) genes derived from flax. The associated SAD promoter sequences are useful for the modification of flax and other plants for the expression of endogenous or foreign genes. However the methods taught by WO 98/18948 are limited by the fact that the SAD promoters are not seed-specific in flax and confer expression to leaves, stems, flowers and seeds. Expression of non-native genes thus may result in undesirable side effects in non-seed tissues. In addition the use of the SAD promoters allows limited control over expression level and timing of expression.

There is a need in the art to further improve methods for the expression of non-native genes in flax seeds and other plant seeds.

SUMMARY OF THE INVENTION

The present invention relates to improved methods for the seed-specific expression of non-native genes in plants. In particular, the invention relates to improved methods for the seed-specific expression of non-native genes in flax.

Accordingly, in one aspect, the invention provides a method for the expression of a nucleic acid sequence of interest in plant seeds comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components

(1) a seed-specific promoter obtained from flax; and

(2) the nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said flax seed-specific promoter;

(b) introducing said chimeric nucleic acid construct into a plant cell; and

(c) growing said flax plant cell into a mature plant capable of setting seed, wherein said nucleic acid sequence of interest is expressed in the seed under the control of said flax seed-specific promoter.

In a preferred embodiment the seed-specific promoter obtained from flax comprises at least one of the following promoter elements: an RY-repeat, an ABRE, an E-BOX or an SEF3 element. Preferably the seed specific promoter obtained from flax comprises an RY-repeat and an ABRE.

In a further preferred embodiment of the invention, at least one expression characteristic, e.g. timing of expression in the plant's life cycle, conferred by the promoter to the non-native nucleic acid sequence is similar to that expression characteristic when conferred to a native nucleic acid sequence. In further preferred embodiments, the flax seed-specific promoter is an oleosin promoter, a 2S storage protein promoter or a legumin-like seed storage protein promoter.

In a further aspect, the present invention provides transgenic seeds prepared according to a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

- (1) a seed-specific promoter obtained from flax; and
- (2) a nucleic acid sequence of interest wherein said nucleic acid
of interest is non-native to said seed-specific
promoter;

(b) introducing said chimeric nucleic acid construct into a plant
cell; and

(c) growing said plant cell into a mature plant capable of
setting seed, wherein said nucleic acid sequence of interest is expressed
in the seed under the control of said seed-specific promoter.

In a further aspect the present invention provides plants capable
of setting seed prepared by a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in
the 5' to 3' direction of transcription as operably linked components:

- (1) a seed-specific promoter obtained from flax; and
- (2) a nucleic acid sequence of interest wherein said nucleic acid
of interest is non-native to said seed-specific
promoter;

(b) introducing said chimeric nucleic acid construct into a plant
cell; and

(c) growing said plant cell into a mature plant capable of
setting seed, wherein said nucleic acid sequence of interest is expressed
in the seed under the control of said seed-specific promoter.

In yet a further aspect, the present invention provides novel flax
seed specific promoters useful for the expression of non-native genes
in flax seeds and the seeds of other plant species, useful for example for
modification of the protein or oil composition of the seed. In a
preferred embodiment the seed-specific promoter obtained from flax
comprises at least one of the following promoter elements: an RY-
repeat, an ABRE, an E-BOX or an SEF3 element. Preferably the seed
specific promoter obtained from flax comprises and RY-repeat and an
ABRE.

In a specific embodiment, the seed specific promoter comprises:

(a) a nucleic acid sequence as shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

5 (b) a nucleic acid sequence that is complimentary to a nucleic acid sequence of (a);

(c) a nucleic acid sequence that has substantial sequence homology to a nucleic acid sequence of (a) or (b);

(d) a nucleic acid sequence that is an analog of a nucleic acid sequence of (a), (b) or (c); or

10 (e) a nucleic acid sequence that hybridizes to a nucleic acid sequence of (a), (b), (c) or (d) under stringent hybridization conditions.

In another aspect, the invention provides chimeric nucleic acid sequences comprising a first nucleic acid sequence obtained from flax operatively linked to a second nucleic acid sequence non-native to said first nucleic acid sequence wherein said first nucleic acid sequence comprises a novel flax seed-specific promoter.

15 Other features and advantages of the present invention will become readily apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art of this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

25 The invention will now be described in relation to the drawings in which:

Figure 1 shows the DNA sequence (SEQ.ID.NO.:1) of a flax genomic clone encoding a 16.0 kDa oleosin protein (SEQ.ID.NOS.:2 and 3).

30 Figure 2 shows the DNA sequence (SEQ.ID.NO.:4) of a flax genomic clone encoding a 18.6 kDa oleosin protein (SEQ.ID.NO.:5).

Figure 3 shows the DNA sequence (SEQ.ID.NO.:6) of a flax genomic clone encoding a 2S storage protein (SEQ.ID.NO.:7).

Figure 4 shows the DNA sequence (SEQ.ID.NO.:8) of a flax genomic clone encoding a 54.5 kDa legumin-like storage protein (SEQ.ID.NOS.:9-12).

5 Figure 5 shows Southern blot analysis of flax genomic DNA probed with flax oleosin DNA sequences.

Figure 6 shows a Northern blot analysis of the seed specific expression of flax oleosins.

Figure 7 shows a Northern blot analysis of the developmental expression of flax oleosins during seed development.

10 Figure 8 shows the GUS activity of flax embryos bombarded with flax oleosin promoter-GUS-flax terminator gene constructs.

Figure 9 shows GUS expression in developing flax embryos and Arabidopsis seeds of plants transformed with a 2S protein gene promoter GUS fusion.

15 Figure 10 shows the tissue-specific expression of GUS in transgenic flax plants transformed with a linin promoter-GUS-linin terminator gene construct.

Figure 11 shows the temporal expression of GUS in transgenic flax plants transformed a linin promoter-GUS-linin terminator gene construct.

20 Figure 12 shows the expression of GUS in transgenic *Brassica napus* plants (L1 to L9) transformed with a linin promoter-GUS-linin terminator gene construct.

Figure 13 shows the expression of GUS in transgenic Arabidopsis plants transformed with a linin promoter-GUS-linin terminator gene construct at different stages of seed development.

DETAILED DESCRIPTION OF THE INVENTION

I. Seed-Specific Promoters

30 The present invention provides novel flax seed specific promoters useful for the expression of non-native genes in flax seeds and the seeds of other plant species. The promoters may be used to modify for example the protein, oil or polysaccharide composition of the seeds.

35 The term "seed-specific promoter", means that a gene expressed under the control of the promoter is predominantly

expressed in plant seeds with no or no substantial expression, typically less than 5% of the overall expression level, in other plant tissues.

In one aspect, the invention provides a flax seed specific promoter which comprises one or more of the following promoter elements purine-pyrimidine sequences (RY repeats), abscisic acid-responsive elements (ABRE), E-Box elements (CANNTG element) and soybean embryo factor 3 elements (SEF3).

As used herein, the RY promoter element is defined as a promoter element having a sequence of CATGCA in accordance with the consensus sequence for RY promoter elements as discussed by Dickenson *et al.* (1988) Nucl. Acids Res. 16:371. The RY elements are found in seed-specific promoters of both monocots and dicots including the *napin* genes of *Brassica napus* (Ellerstrom *et al.* (1996) Plant Mol. Biol. 32:1019), and the maize *C1* gene (Kao *et al.* (1996) Plant Cell 8:1171). It should be noted that RY elements have also been referred to as the "legumin box", the "Sph" element or the "G-box".

The phytohormone abscisic acid (ABA) is necessary for regulation of several events during late seed development (Busk and Pages. (1998) Plant Mol. Biol. 37:425-435.) As used herein, the abscisic acid responsive element (ABRE) is defined as a promoter element having a sequence of 8-10 base pairs with a core sequence of ACGT in accordance with the consensus sequence for ABRE as defined by Izawa *et al.* (1993) J. Mol. Biol. 14:61-72 and Shen and Ho (1995) Plant Cell 7:295-307. Hobo *et al.* (1999) PNAS 96:15348-15353 have performed promoter dissection studies and further identified the ABRE sequence to be (G/C/T)ACGT(G/T)GC. Expression studies have also shown that the sequence (C/T)ACGTGGC is a strong ABRE (Busk *et al.* (1999) Plant Mol Biol 41:529-536 and Shen and Ho (1995) Plant Cell 7:295-307). The ABRE elements have been identified in many seed-specific promoters including *Em*, a class I Late Embryogenesis Abundant protein in Arabidopsis (Carles *et al.* (2002) Plant J. 30(3):373-83.), *rab17* gene from maize (Busk *et al.* (1997) Plant J. 11(6):1285-95) and the *Brassica napus* storage protein gene *napA* (Ellerstrom *et al.* (1996) Plant Mol Biol. 32(6):1019-27).

As used herein, the E-box motif promoter element is defined as a promoter element which is a basic region helix-loop-helix with the sequence CANNTG in accordance with the consensus sequence for E-box motifs as defined by Yoder *et al.* (1999) *Biochimica et Biophysica Acta* 1446:403-413. The E-box element is found in seed-specific genes including a cotton palmitoyl-acyl carrier protein (ACP) thioesterase (*Fat B1*) (Yoder *et al.* (1999) *Biochimica et Biophysica Acta* 1446:403-413) and has shown to be involved in both tissue specific and developmental control of the *Phaseolus vulgaris* seed storage glycoprotein Phaseolin (Burrow, *et al.* (1992) *Plant J* 2: 537-548 and Kawagoe *et al.* (1994) *Plant J* 5: 885-890).

Finally the soybean embryo factor 3 (SEF3) responsive element has a consensus of AACCCA (Allen *et al.* (1989). *Plant Cell* 1: 623-631). The SEF3 element has been identified in the promoter of the α' subunit of the soybean α' -conglycinin 7S seed storage protein (Allen *et al.* (1989) *Plant Cell* 1: 623-631).

In a preferred embodiment, the seed-specific promoter obtained from flax comprises at least one of the following promoter elements: an RY-repeat, an ABRE, and E-Box or an SEF3 element. In accordance with the present invention, the flax seed specific promoter preferably comprises an RY repeat and an ABRE. Preferably, the RY repeat and the ABRE are separated from each other by less than 100 nucleotides and more than 20 nucleotides. More preferably, the RY repeat and the ABRE are separated from each other by less than 50 nucleotides and more than 20 nucleotides and most preferably, the RY element and the ABRE are separated from each other less than 35 nucleotides and more than 20 nucleotides.

In particular, the inventors have isolated and sequenced four seed-specific promoters from flax. The four promoters share structural similarities in that they all have an alternating purine-pyrimidine sequences (RY repeat) and abscisic acid-responsive element (ABRE) with similar spacing. In addition, the 2S and linin promoters both have the following four promoter elements: EBOX (CANNTG element), abscisic acid-responsive element (ABRE), soybean embryo factor 3 (SEF3), alternating purine-pyrimidine sequences (RY repeat) and the

order of the elements is maintained. The two oleosin promoters share the promoter elements RY repeat and CACGTG sequence (which has ABRE as a sub-set) and the order of the elements is also maintained. The elements found in each of the four promoters are listed in the table below.

5

| Specification Figure | Promoter | Element | Nucleotides |
|----------------------|-----------------------------|--------------|----------------------------|
| 1 | 16 kDa oleosin promoter | RY repeat | • 1817-1822 • 1821-1826 |
| | | ABRE element | • 1880-1884 |
| 2 | 18.6 kDa oleosin promoter | RY repeat | • 1677-1682 |
| | | ABRE element | • 1711-1750 |
| 3 | 2S Storage protein promoter | RY repeat | • 285-290 |
| | | ABRE element | • 253-257 |
| | | EBOX | • 234-239 |
| | | SEF3 | • 269-274 |
| 4 | Linin promoter | RY repeat | • 1888-1893 |
| | | ABRE element | • 1860-1864 |
| | | EBOX | • 1836-1841 |
| | | SEF3 | • 1877-1882 |

In one embodiment the flax seed specific promoter of the present invention, in addition to the RY-repeat and the ABRE, comprises an EBOX element and a SEF3 element. The EBOX element is preferably located between 250 and 150 nucleotides from the "ATG" start codon. More preferably, the EBOX element is located between 200 and 175 nucleotides from the start codon and most preferably, the EBOX element is located are between 195 and 185 nucleotides, inclusive, from the "ATG" start codon. Preferably, the SEF3 element is located between 200 and 50 nucleotides from the "ATG" start codon. More preferably, the SEF3 element is located between 175 and 100 nucleotides from the start codon and most preferably, the SEF3 element is located between 160 and 145 nucleotides, inclusive, from the "ATG" start codon.

In a preferred embodiment, the seed specific promoter comprises:

10

15

20

(a) a nucleic acid sequence as shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

5 (b) a nucleic acid sequence that is complimentary to a nucleic acid sequence of (a);

(c) a nucleic acid sequence that has substantial sequence homology to a nucleic acid sequence of (a) or (b);

(d) a nucleic acid sequence that is an analog of a nucleic acid sequence of (a), (b) or (c); or

10 (e) a nucleic acid sequence that hybridizes to a nucleic acid sequence of (a), (b), (c) or (d) under stringent hybridization conditions.

The term "sequence that has substantial sequence homology" means those nucleic acid sequences which have slight or inconsequential sequence variations from the sequences in (a) or (b),
15 i.e., the sequences function in substantially the same manner and are capable of driving seed specific expression of non-native nucleic acid sequences. The variations may be attributable to local mutations or structural modifications. Nucleic acid sequences having substantial homology include nucleic acid sequences having at least 65%, more
20 preferably at least 85%, and most preferably 90-95% identity with the nucleic acid sequences as shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or Figure 4 (SEQ.ID.NO.:8). Sequence identity can be calculated according to methods known in the art. Sequence identity is most preferably assessed by the algorithm of
25 BLAST version 2.1 advanced search. BLAST is a series of programs that are available online at <http://www.ncbi.nlm.nih.gov/BLAST>. The advanced blast search (<http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=1>) is set to default parameters. (ie Matrix BLOSUM62; Gap existence cost 11; Per
30 residue gap cost 1; Lambda ratio 0.85 default). References to BLAST searches are: Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403410; Gish, W. & States, D.J. (1993) "Identification of protein coding regions by database similarity search." Nature Genet. 3:266272; Madden, T.L.,
35 Tatusov, R.L. & Zhang, J. (1996) "Applications of network BLAST

server" Meth. Enzymol. 266:131_141; Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI_BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:33893402; Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649656.

The term "sequence that hybridizes" means a nucleic acid sequence that can hybridize to a sequence of (a), (b), (c) or (d) under stringent hybridization conditions. Appropriate "stringent hybridization conditions" which promote DNA hybridization are known to those skilled in the art, or may be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. The term "stringent hybridization conditions" as used herein means that conditions are selected which promote selective hybridization between two complementary nucleic acid molecules in solution. Hybridization may occur to all or a portion of a nucleic acid sequence molecule. The hybridizing portion is at least 50% the length with respect to one of the polynucleotide sequences encoding a polypeptide. In this regard, the stability of a nucleic acid duplex, or hybrids, is determined by the T_m , which in sodium containing buffers is a function of the sodium ion concentration, G/C content of labeled nucleic acid, length of nucleic acid probe (l), and temperature ($T_m = 81.5^{\circ}\text{C} - 16.6 (\text{Log}_{10} [\text{Na}^+]) + 0.41(\%(G+C) - 600/l)$). Accordingly, the parameters in the wash conditions that determine hybrid stability are sodium ion concentration and temperature. In order to identify molecules that are similar, but not identical, to a known nucleic acid molecule a 1% mismatch may be assumed to result in about a 1°C decrease in T_m , for example if nucleic acid molecules are sought that have a greater than 95% identity, the final wash will be reduced by 5°C . Based on these considerations stringent hybridization conditions shall be defined as: hybridization at 5 x sodium chloride/sodium citrate (SSC)/5 x Denhardt's solution/1.0% SDS at T_m (based on the above equation) - 5°C , followed by a wash of 0.2 x SSC/0.1% SDS at 60°C .

The term "a nucleic acid sequence which is an analog" means a nucleic acid sequence which has been modified as compared to the sequence of (a), (b) or (c) wherein the modification does not alter the utility of the sequence (i.e. as a seed specific promoter) as described herein. The modified sequence or analog may have improved properties over the sequence shown in (a), (b) or (c). One example of a modification to prepare an analog is to replace one of the naturally occurring bases (i.e. adenine, guanine, cytosine or thymidine) of the sequence shown in Figure 1 (SEQ ID NO:1), Figure 2 (SEQ ID NO:4), Figure 3 (SEQ ID NO:6) or Figure 4 (SEQ ID NO:8) with a modified base such as such as xanthine, hypoxanthine, 2-aminoadenine, 6-methyl, 2-propyl and other alkyl adenines, 5-halo uracil, 5-halo cytosine, 6-aza uracil, 6-aza cytosine and 6-aza thymine, pseudo uracil, 4-thiouracil, 8-halo adenine, 8-aminoadenine, 8-thiol adenine, 8-thiolalkyl adenines, 8-hydroxyl adenine and other 8-substituted adenines, 8-halo guanines, 8 amino guanine, 8-thiol guanine, 8-thiolalkyl guanines, 8-hydroxyl guanine and other 8-substituted guanines, other aza and deaza uracils, thymidines, cytosines, adenines, or guanines, 5-trifluoromethyl uracil and 5-trifluoro cytosine.

Another example of a modification is to include modified phosphorous or oxygen heteroatoms in the phosphate backbone, short chain alkyl or cycloalkyl intersugar linkages or short chain heteroatomic or heterocyclic intersugar linkages in the nucleic acid molecule shown in Figure 1 (SEQ ID NO:1), Figure 2 (SEQ ID NO:4), Figure 3 (SEQ ID NO:6) or Figure 4 (SEQ ID NO:8). For example, the nucleic acid sequences may contain phosphorothioates, phosphotriesters, methyl phosphonates, and phosphorodithioates.

A further example of an analog of a nucleic acid molecule of the invention is a peptide nucleic acid (PNA) wherein the deoxyribose (or ribose) phosphate backbone in the DNA (or RNA), is replaced with a polyamide backbone which is similar to that found in peptides (P.E. Nielsen, et al Science 1991, 254, 1497). PNA analogs have been shown to be resistant to degradation by enzymes and to have extended lives *in vivo* and *in vitro*. PNAs also bind stronger to a complimentary DNA sequence due to the lack of charge repulsion between the PNA strand

and the DNA strand. Other nucleic acid analogs may contain nucleotides containing polymer backbones, cyclic backbones, or acyclic backbones. For example, the nucleotides may have morpholino backbone structures (U.S. Pat. No. 5,034,506). The analogs may also
5 contain groups such as reporter groups, a group for improving the pharmacokinetic or pharmacodynamic properties of nucleic acid sequence.

In a specific embodiment, the promoter is selected from the group of promoters consisting of Figure 1 (SEQ ID NO:1), Figure 2
10 (SEQ ID NO:4), Figure 3 (SEQ ID NO:6) and Figure 4 (SEQ ID NO:8) or a nucleic acid sequence hybridizing thereto under stringent conditions.

In another aspect, the invention provides chimeric nucleic acid sequences comprising a first nucleic acid sequence obtained from flax operatively linked to a second nucleic acid sequence non-native to said
15 first nucleic acid sequence wherein said first nucleic acid sequence comprises a novel flax seed-specific promoter.

In accordance with the present invention, the chimeric nucleic acid sequences can be incorporated in a known manner in a recombinant expression vector which ensures good expression in the
20 seed cell. Accordingly, the present invention includes a recombinant expression vector comprising a chimeric nucleic acid sequence of the present invention suitable for expression in a seed cell.

The term "suitable for expression in a seed cell" means that the recombinant expression vectors contain the chimeric nucleic acids
25 sequence of the invention, a regulatory region and a termination region, selected on the basis of the seed cell to be used for expression, which is operatively linked to the nucleic acid sequence encoding the polypeptide of desirable amino acid composition. Operatively linked is intended to mean that the chimeric nucleic acid sequence encoding the
30 polypeptide is linked to a regulatory sequence and termination region which allows expression in the seed cell. A typical construct consists, in the 5' to 3' direction of a regulatory region complete with a promoter capable of directing expression in a plant, a polypeptide coding region and a transcription termination region functional in plant cells. These
35 constructs may be prepared in accordance with methodology well

known to those of skill in the art of molecular biology (see for example: Sambrook *et al.* (1990), Molecular Cloning, 2nd ed. Cold Spring Harbor Press). The preparation of constructs may involve techniques such as restriction digestion, ligation, gel electrophoresis, DNA sequencing and PCR. A wide variety of cloning vectors is available to perform the necessary cloning steps. Especially suitable for this purpose are the cloning vectors with a replication system that is functional in *Escherichia coli* such as pBR322, the pUC series M13mp series, pACYC184, pBluescript etc. Nucleic acid sequences may be introduced into these vectors and the vectors may be used to transform *E. coli* which may be grown in an appropriate medium. Plasmids may be recovered from the cells upon harvesting and lysing the cells. Final constructs may be introduced into plant vectors compatible with integration into the plant such as the Ti and Ri plasmids.

II. Methods

As hereinbefore mentioned, the present invention relates to improved methods for the expression of non-native genes in plants, in particular flax. The invention provides methods allowing the seed-specific expression of non-native genes in flax and other plants. The methods of the invention are advantageous in that improved control over the expression of non-native genes in seeds is obtained. Expression of the non-native gene is restricted to the seeds, thereby limiting potential undesirable effects resulting from the expression in other plant organs or tissues. In addition, the provided methodology allows improved control over expression characteristics, such as the expression level of the non-native gene and timing of expression of the non-native gene in the developmental cycle of the plant. The methods of the present invention are particularly useful in that in accordance with the present invention the seed composition with respect to valuable raw materials, such as oil, protein and polysaccharides, may be altered both qualitatively and quantitatively.

Accordingly, in one aspect, the invention provides a method for the expression of a nucleic acid sequence of interest in plant seeds comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components;

(1) a seed-specific promoter obtained from flax; and

5 (2) the nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said flax seed-specific promoter;

(b) introducing said chimeric nucleic acid construct into a plant cell; and

10 (c) growing said plant cell into a mature plant capable of setting seed, wherein said nucleic acid sequence of interest is expressed in the seed under the control of said flax seed-specific promoter.

As used herein, the term "non-native" refers to any nucleic acid sequence, including any RNA or DNA sequence, which is not normally associated with the seed-specific promoter. This includes heterologous
15 nucleic acid sequences which are obtained from a different plant species as the promoter as well as homologous nucleic acid sequences which are obtained from the same plant species as the promoter but are not associated with the promoter in the wild-type (non-transgenic) plant.

20 The non-native nucleic acid sequence when linked to a seed-specific promoter obtained from flax results in a chimeric construct. The chimeric construct is introduced into a flax plant cell to create a transgenic flax plant cell which results in a detectably different phenotype of the flax plant cell or flax plant grown from it when
25 compared with a non-transgenic flax plant cell or flax plant grown from it. A contiguous nucleic acid sequence identical to the nucleic acid sequence of the chimeric construct is not present in the non-transformed flax plant cell or flax plant grown from it. In this respect, chimeric nucleic acid sequences include those sequences which contain
30 a flax promoter linked to a nucleic acid sequence obtained from another plant species or a nucleic acid sequence from flax but normally not associated with that promoter. Chimeric nucleic acid sequences as used herein further include sequences comprising a flax promoter and a nucleic acid sequence that is normally linked to the promoter but
35 additionally containing a non-native nucleic acid sequence. For

example, if the promoter is a flax seed-specific oleosin promoter, sequences "non-native" to the flax oleosin promoter also include a sequence comprising a fusion between the flax oleosin gene naturally associated with the oleosin promoter, and a coding sequence of interest that is not naturally associated with the promoter. The term non-native is also meant to include a fusion gene as hereinabove which additionally includes a cleavage sequence separating the nucleic acid sequence that is normally linked to the promoter sequence and the gene encoding the protein of interest.

The term "nucleic acid sequence" refers to a sequence of nucleotide or nucleoside monomers consisting of naturally occurring bases, sugars and intersugar (backbone) linkages. The term also includes modified or substituted sequences comprising non-naturally occurring monomers or portions thereof, which function similarly. The nucleic acid sequences of the present invention may be ribonucleic (RNA) or deoxyribonucleic acids (DNA) and may contain naturally occurring bases including adenine, guanine, cytosine, thymidine and uracil. The sequences may also contain modified bases such as xanthine, hypoxanthine, 2-aminoadenine, 6-methyl, 2-propyl, and other alkyl adenines, 5-halo uracil, 5-halo cytosine, 6-aza uracil, 6-aza cytosine and 6-aza thymine, pseudo uracil, 4-thiouracil, 8-halo adenine, 8-amino adenine, 8-thiol adenine, 8-thio-alkyl adenines, 8-hydroxyl adenine and other 8-substituted adenines, 8-halo guanines, 8-amino guanine, 8-thiol guanine, 8-thioalkyl guanines, 8-hydroxyl guanine and other 8-substituted guanines, other aza and deaza uracils, thymidines, cytosines, adenines, or guanines, 5-trifluoromethyl uracil and 5-trifluoro cytosine. The methods for the expression of non-native genes in flax seeds in accordance with the present invention may be practiced using any flax seed-specific promoter and are not limited by the specific flax seed specific promoter that is selected. Preferred flax seed-specific promoters are described in Section I. In preferred embodiments of the present invention, the flax seed-specific promoter confers to the non-native nucleic acid sequence at least one expression characteristic which is similar or identical to an expression characteristic conferred to the native nucleic acid sequence by the native promoter.

The term "expression characteristic" as used herein refers to any measurable property or effect conferred by the flax seed-specific promoter to the nucleic acid sequence operably linked to the flax seed-specific promoter. Thus in preferred embodiments, timing of expression in the plant's life cycle, of the non-native nucleic acid sequence is similar or identical to timing of expression of the native nucleic acid sequence. In further preferred embodiments, the expression level of the heterologous nucleic acid sequence is similar or identical to the expression level of the native nucleic acid sequence. In yet further specific embodiments, the response of the non-native gene to alterations in lighting conditions, changes in wavelength or light intensity for example, changes in temperature, tissue wounding, changes in concentration of chemical agents, such as for example phytohormones and pesticides, is similar to the response of the native nucleic acid sequence to these stimuli. Other desired expression characteristics conferred by a flax seed-specific promoter may be recognized by those skilled in the art and a flax seed-specific promoter may be selected accordingly.

Flax-seed specific promoters that may be used in accordance with the present invention include promoters associated with seed storage proteins, such as all albumins and globulins, including the vicilin and legumin-like proteins, as well as seed-specific promoters not associated with seed storage proteins, such as oleosins. Of further particular interest are promoters associated with fatty acid metabolism, such as acyl carrier protein (ACP), saturases, desaturases, elongases and the like.

The flax seed-specific promoter preferably comprises one or more of the following promoter elements purine-pyrimidine sequences (RY repeats), abscisic acid-responsive elements (ABRE), E-Box elements (CANNTG element) and soybean embryo factor 3 elements (SEF3). In preferred embodiments of the present invention the seed specific promoter used is an oleosin promoter, a legumin-like seed storage protein promoter or a 2S storage protein promoter. In particularly preferred embodiments the seed specific promoter has the sequence shown in Figure 1 (SEQ ID NO:1), Figure 2 (SEQ ID NO:4),

Figure 3 (SEQ ID NO:6) or Figure 4 (SEQ ID NO:8) or any nucleic acid sequences obtainable from flax and hybridizing to any one of these four nucleic acid sequences under stringent conditions.

The present invention further provides methods of use for the novel promoters shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) and Figure 4 (SEQ.ID.NO.:8) in all plant species including flax. Accordingly, the invention also includes the preparation of chimeric nucleic acid constructs comprising a promoter selected from the group promoters shown in Figure 1 (SEQ ID NO:1), Figure 2 (SEQ ID NO:4), Figure 3 (SEQ ID NO:6) and Figure 4 (SEQ ID NO:8) and a nucleic acid sequence of interest, and expression in a seed-specific manner of the nucleic acid sequence of interest in a plant species such as flax and under the control of the flax promoter.

Accordingly, in a specific embodiment there is provided a method for the expression of a nucleic acid sequence of interest in plant seeds comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components;

(1) a seed-specific promoter selected from the group of seed-specific promoters consisting of

(i) a nucleic acid sequence as shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) or Figure 4 (SEQ.ID.NO.:8) wherein T can also be U;

(ii) a nucleic acid sequence that is complimentary to a nucleic acid sequence of (i);

(iii) a nucleic acid sequence that has substantial sequence homology to a nucleic acid sequence of (i) or (ii); and

(iv) a nucleic acid sequence that is an analog of a nucleic acid sequence of (i), (ii) or (iii);

(v) a nucleic acid sequence that hybridizes to a nucleic acid sequence of (i), (ii), (iii) or (iv) under stringent hybridization conditions; and

(2) said nucleic acid of interest;

(b) introducing the chimeric nucleic acid construct into a plant cell;

(c) growing said plant cell into a mature plant capable of setting seed, wherein said nucleic acid sequence of interest is expressed
5 in the seed under the control of said seed-specific promoter.

Additional flax seed-specific promoters may be used in accordance with the present invention. These promoters may be obtained in a number of ways. Where a flax seed protein has been isolated, it may be partially sequenced, so that a nucleic acid probe may
10 be designed for identifying RNA specific to the seed. To further enhance the RNA specifically associated with the seed, cDNA may be prepared from seed cells and the cDNA may be subtracted with mRNA or cDNA from non-seed cells. The remaining seed cDNA may then be used to probe a genomic DNA library for complementary
15 sequences. Sequences hybridizing to the cDNA may subsequently be obtained and the associated promoter region may be isolated. It is also possible to screen genomic DNA libraries prepared from flax seed tissues using known seed specific genes from other plant species and subsequently isolate their associated promoters. Due to the relative
20 abundance of seed-storage proteins in seeds it is also possible to obtain sequence information through random sequencing of flax seed cDNA libraries. Those cDNA sequences matching sequence of known seed-storage proteins could be used to identify the associated promoter. Databases containing sequence information from large
25 scale sequencing from for example *Arabidopsis* and maize may be searched for known seed-specific proteins and/or promoters and the information may be used to identify promoter sequences in flax that share sequence similarity. Alternative methods to isolate additional flax seed specific promoters may be used and novel flax seed specific
30 promoters may be discovered by those skilled in the art and used in accordance with the present invention.

The nucleic acid sequence of interest linked to the promoter may be any nucleic acid sequence of interest including any RNA or DNA sequence encoding a peptide or protein of interest, for example,
35 an enzyme, or a sequence complementary to a genomic sequence,

where the genomic sequence may be at least one of an open reading frame, an intron, a non-coding leader sequence, or any sequence where the complementary sequence will inhibit transcription, messenger RNA processing, for example splicing or translation. The nucleic acid sequence of interest may be synthetic, naturally derived or a combination thereof. As well, the nucleic acid sequence of interest could be a fragment of the natural sequence, for example just include the catalytic domain or a structure of particular importance. Depending upon the nature of the nucleic acid sequence of interest, it may be desirable to synthesize the sequence with plant preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in particular plant species of interest.

The nucleic acid sequence of interest may encode any of a variety of recombinant proteins. Examples of recombinant proteins which might be expressed by the methods of the present invention include proteins with a favorable catalytic function or a valuable protein that will accumulate to high levels and be extracted if desired. Proteins with a catalytic function, include, but are not limited to, proteins that confer a new biochemical phenotype on the developing seeds. New phenotypes could include such modifications as altered seed-protein or seed oil composition or seed polysaccharide composition, enhanced production of pre-existing desirable products or properties and the reduction or even suppression on an undesirable gene product using antisense, ribozyme or co-suppression technologies (Izant and Weintraub (1984) Cell 26: 1007-1015, antisense; Hazelhoff and Gerlach (1988) Nature 334: 585-591, ribozyme; Napoli *et al.* (1990) Plant Cell 2: 279-289, co-suppression).

It is expected that the desired proteins would be expressed in all embryonic tissues, although varying cellular expression may be detected in the different embryonic tissues such as the embryonic axis and cotyledons. The nucleic acid sequence of interest may be expressed at any stage in seed development. The timing of expression may depend on the particular use of the invention. Expression of enzymes involved in oil modification may be desirable early in seed

development, for example before accumulation of seed storage protein.

Besides the promoter region and the nucleic acid sequence of interest, a nucleic acid sequence capable of terminating transcription is typically included in expression vectors. Transcriptional terminators are preferably about 200 to about 1,000 nucleotide base pairs and may comprise any such sequences functional in plants, such as the nopaline synthase termination region (Bevan *et al.*, (1983) Nucl. Acid. Res. 11: 369-385), the phaseolin terminator (van der Geest *et al.*, (1994) Plant J. 6(3): 413-423), the terminator for the octopine synthase gene of *Agrobacterium tumefaciens* or other similarly functioning elements. These transcription terminator regions can be obtained as described by An (1987), Methods in Enzym. 153: 292 or are already present in plasmids available from commercial sources such as ClonTech, Palo Alto, California. The choice of the appropriate terminator may have an effect of the rate of transcription.

The chimeric construct may further comprise enhancers such as the AMV leader (Jobling and Gehrke (1987), Nature 325: 622-625) or introns. It should be understood that the design of the expression vector may depend on such factors as the choice of the plant species and/or the type of polypeptide to be expressed.

The expression vectors will normally also contain a marker gene. Marker genes comprise all genes that enable distinction of transformed plant cells from non-transformed cells, including selectable and screenable marker genes. Conveniently, a marker may be a resistance marker to a herbicide, for example, glyphosate or phosphinothricin, or to an antibiotic such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol and the like, which confer a trait that can be selected for by chemical means. Screenable markers may be employed to identify transformants through observation. They include but are not limited to the β -glucuronidase or *uidA* gene, a β -lactamase gene or a green fluorescent protein (Niedz *et al.* (1995) Plant Cell Rep. 14: 403).

In order to introduce nucleic acid sequences into plant cells in general a variety of techniques are available to the skilled artisan.

Agrobacterium-mediated transformation for flax plant cells has been reported and flax transformants may be obtained in accordance with the methods taught by Dong and McHughen (1993) Plant Science 88: 61-77, although a variety of other techniques (see below) may also be used to introduce the chimeric DNA constructs in flax cells if so desired.

Transformed flax plants grown in accordance with conventional agricultural practices known to a person skilled in the art are allowed to set seed. Flax seed may then be obtained from mature flax plants and analyzed for desired altered properties with respect to the wild-type seed.

Two or more generations of plants may be grown and either crossed or selfed to allow identification of plants and strains with desired phenotypic characteristics including production of the recombinant polypeptide. It may be desirable to ensure homozygosity in the plants to assure continued inheritance of the recombinant trait. Methods for selecting homozygous plants are well known to those skilled in the art of plant breeding and include recurrent selfing and selection and anther and microspore culture. Homozygous plants may also be obtained by transformation of haploid cells or tissues followed by regeneration of haploid plantlets subsequently converted to diploid plants by any number of known means (e.g. treatment with colchicine or other microtubule disrupting agents).

The present invention also includes transgenic plant seeds prepared according to a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

- (1) a seed-specific promoter obtained from flax; and
- (2) a nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said seed-specific promoter;

(b) introducing said chimeric nucleic acid construct into a plant cell; and

(c) growing said plant cell into a mature plant capable of setting seed

wherein said nucleic acid sequence of interest is expressed in the seed under the control of said seed-specific promoter.

The "seed-specific promoter obtained from flax" is as defined previously in Section I. In preferred embodiments of the invention the seed-specific promoter is selected from the group of flax seed specific promoters consisting of, a 2S storage protein promoter, a globulin promoter, an oleosin promoter, and a legumin-like seed storage protein promoter. Specific promoter sequences are shown in Figure 1 (SEQ.ID.NO.:1), Figure 2 (SEQ.ID.NO.:4), Figure 3 (SEQ.ID.NO.:6) and Figure 4 (SEQ.ID.NO.:8).

The present invention further provides plants capable of setting seed prepared by a method comprising:

(a) preparing a chimeric nucleic acid construct comprising in the 5' to 3' direction of transcription as operably linked components:

- (1) a seed-specific promoter obtained from flax; and
- (2) a nucleic acid sequence of interest wherein said nucleic acid of interest is non-native to said seed-specific promoter;

(b) introducing said chimeric nucleic acid construct into a plant cell; and

(c) growing said flax plant cell into a mature plant capable of setting seed

wherein said nucleic acid sequence of interest is expressed in the seed under the control of said seed-specific promoter.

A variety of techniques are available for the introduction of nucleic acid sequences, in particular DNA, into plant host cells in general. For example, the chimeric DNA constructs may be introduced into host cells obtained from dicotyledonous plants, such as tobacco, and oleoagenous species, such as *Brassica napus* using standard *Agrobacterium* vectors by a transformation protocol such as described by Moloney *et al.* (1989), Plant Cell Rep. 8: 238-242 or Hinchey *et al.* (1988) Bio/Technol. 6: 915-922; or other techniques known to those skilled in the art. For example, the use of T-DNA for transformation of plant cells has received extensive study and is amply described in EP 0 120 516, Hoekema *et al.*, (1985), Chapter V In: *The Binary Plant Vector*

System Offset-drukkerij Kanters BV, Alblasserdam); Knauf *et al.* (1983),
Genetic Analysis of Host Expression by Agrobacterium, p. 245, In: *Molecular*
Genetics of Bacteria-Plant Interaction, Puhler, A. ed. Springer-Verlag,
NY); and An *et al.*, (1985), (EMBO J., 4: 277-284). *Agrobacterium*
5 transformation may also be used to transform monocot plant species
(US Patent 5,591,616).

Conveniently, explants may be cultivated with *Agrobacterium*
tumefaciens or *Agrobacterium rhizogenes* to allow for the transfer of the
transcription construct in the plant host cell. Following transformation
10 using *Agrobacterium* the plant cells are dispersed into an appropriate
medium for selection, subsequently callus, shoots and eventually
plants are recovered. The *Agrobacterium* host will harbour a plasmid
comprising the vir genes necessary for transfer of the T-DNA to plant
cells. For injection and electroporation (see below) disarmed Ti-
15 plasmids (lacking the tumour genes, particularly the T-DNA region)
may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of
constructs described herein to obtain transformation and expression in
a wide variety of monocotyledonous and dicotyledonous plant species.
20 These techniques are especially useful for transformation of plant
species that are intractable in an *Agrobacterium* transformation system.
Other techniques for gene transfer include particle bombardment
(Sanford, (1988), Trends in Biotechn. 6: 299-302), electroporation
(Fromm *et al.*, (1985), PNAS USA, 82: 5824-5828; Riggs and Bates,
25 (1986), PNAS USA 83: 5602-5606), PEG mediated DNA uptake
(Potrykus *et al.*, (1985), Mol. Gen. Genetics., 199: 169-177),
microinjection (Reich *et al.*, Bio/Techn. (1986) 4:1001-1004) and silicone
carbide whiskers (Kaeppeler *et al.* (1990) Plant Cell Rep. 9: 415-418).

In a further specific applications such as to *B. napus*, the host cells
30 targeted to receive recombinant DNA constructs typically will be
derived from cotyledonary petioles as described by Moloney *et al.*
(1989) Plant Cell Rep. 8: 238-242. Other examples using commercial oil
seeds include cotyledon transformation in soybean explants (Hinchee
et al., (1988) Bio/Technol. 6: 915-922) and stem transformation of cotton
35 (Umbeck *et al.*, (1987) Bio/Technol. 5: 263-266).

Following transformation, the cells, for example as leaf discs, are grown in selective medium. Once the shoots begin to emerge, they are excised and placed onto rooting medium. After sufficient roots have formed, the plants are transferred to soil. Putative transformed plants are then tested for presence of a marker. Southern blotting is performed on genomic DNA using an appropriate probe, to show integration into the genome of the host cell.

The methods provided by the present invention can be used in conjunction a broad range of plant species. Particularly preferred plant cells employed in accordance with the present invention include cells from the following plants: soybean (*Glycine max*), rapeseed (*Brassica napus*, *Brassica campestris*), sunflower (*Helianthus annuus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), tobacco (*Nicotiana tabacum*), alfalfa (*Medicago sativa*), wheat (*Triticum* sp.), barley (*Hordeum vulgare*), oats (*Avena sativa* L.), sorghum (*Sorghum bicolor*), *Arabidopsis thaliana*, potato (*Solanum* sp.), flax/linseed (*Linum usitatissimum*), safflower (*Carthamus tinctorius*), oil palm (*Eleais guineeis*), groundnut (*Arachis hypogaea*), Brazil nut (*Bertholletia excelsa*) coconut (*Cocus nucifera*), castor (*Ricinus communis*), coriander (*Coriandrum sativum*), squash (*Cucurbita maxima*), jojoba (*Simmondsia chinensis*) and rice (*Oryza sativa*). Preferably the plant is flax/linseed (*Linum usitatissimum*) or safflower (*Carthamus tinctorius*).

The present invention has a variety of uses which include improving the intrinsic value of plant seeds by their accumulation of altered polypeptides or novel recombinant peptides or by the incorporation or elimination or a metabolic step. Use of the invention may result in improved protein quality (for example, increased concentrations or essential or rare amino acids), improved liquid quality by a modification of fatty acid composition, or improved or elevated carbohydrate composition. Examples include the expression of sulfur-rich proteins, such as those found in lupins or brazil nuts in a seed deficient in sulphurous amino acids. Improved protein quality could also be achieved by the expression of a protein or a fragment of a protein that is enriched in essential amino acids including lysine, cysteine, methionine and tryptophan. Alternatively, a fatty acyl

coenzyme A, a transferase enzyme capable of modifying fatty acid ratios in triglycerides (storage lipid) could be expressed. In cases where a recombinant protein is allowed to accumulate in the seed, the protein could also be a peptide which has pharmaceutical or industrial value. In this case the peptide could be extracted from the seed and used in crude or purified form as appropriate for the intended use. As well, the polypeptides that are expressed in the seeds can be fragments or derivatives or the native protein. Pharmaceutically useful proteins may include, but are not limited to, anticoagulants, such as hirudin, antibodies, including monoclonal antibodies and antibody fragments, vaccines, cytokines or growth factors such as bovine growth factor, cholinergic differentiation factor (CDF), ciliary neurotrophic factor (CNTF), fibroblast growth factor (FGF), fish growth factor, gonadotropin, granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), human growth hormone, interferon alpha (IFN- α), interferon beta (IFN- β), interferon gamma (IFN- γ), interleukin 1-alpha (IL1- α), interleukin 1-beta (IL1- β), interleukin-2 (IL-2), interleukin-3 (IL-3), interleukin-4 (IL-4), interleukin-5 (IL-5), interleukin-6 (IL-6), interleukin-10 (IL-10), leukemia inhibitory factor (LIF), thioredoxin, macrophage colony-stimulating factor (M-CSF), myelomonocytic growth factor, nerve growth factor (NGF), oncostatin M, platelet-derived growth factor (PDGF), prolactin, transforming growth factor alpha (TGF- α), transforming growth factor beta2 (TGF- β 2), tumour necrosis factor alpha (TNF- α), and tumour necrosis factor beta (TNF- β). Pharmaceutically useful proteins can also include mammalian proteins, for example, but not limited to α -1-antitrypsin, anti-obesity proteins, blood proteins, collagen, collagenase, elastin, elastase, enteropeptidase, fibrinogen, haemoglobin, human serum albumin, insulin, lactoferrin, myoglobin and pulmonary surfactant proteins.

Industrially useful peptides may include, but are not limited to α -amylase or other amylases, amyloglucosidase, arabinase, catalase, cellobiohydrolase, cellulases, chitinases, chymotrypsin, dehydrogenases, endo-glucanase, chymosin, endo-galactanase,

esterases, β -galactosidase, α -galactosidase or other galactosidases, gastric lipases, glucanases, glucose isomerase, hemi-cellulases, hydrolases, isomerase, ligninases, lipases, lyases, lysozymes, oxidases, oxidoreductase, papain, pectinases, pectin lyase, peroxidases, phosphatases, phytase, proteases, pullulanases, reductases, serine proteases, thioredoxin, transferase, trypsin, and xylanase.

The following non-limiting examples are illustrative of the present invention:

EXAMPLES

EXAMPLE 1

Isolation of Seed-Specific Flax Promoters

Seed specific cDNA clones were isolated from a flax seed specific cDNA-library. These cDNA clones were sequenced and the Basic Local Alignment Search Tool (BLAST) was used to compare these sequences against others in public databases such as Genbank. This comparison revealed that the deduced amino acid sequence of several of the isolated cDNAs had a high degree of similarity to both the low and high molecular weight class of oleosins, 2S-albumin and legumin-like storage proteins. Probes were prepared individually from (portions of) cDNAs encoding oleosins, 2S albumin and legumin-like storage proteins and these were used to screen a genomic library prepared from the flax line Forge that is homozygous for four rust resistance genes (Anderson *et al.* (1997), The Plant Cell 9: 641-651). Several positive lambda clones for each probe were identified after high-stringency screening. The inserts were subcloned into the plasmid vector pBluescript and sequenced. Sequence information revealed that we had isolated the genomic counterparts to the oleosins, 2S albumin and cDNAs *legumin-like* cDNAs. Sequence information of the genomic clones containing sequences encoding a high and low molecular weight oleosin isoforms, 2S albumin and a legumin-like gene are presented in Figures 1 to 4 respectively.

Figure 1 and SEQ.ID.NO.:1 shows the DNA sequence of a flax genomic clone encoding a 16.0 kDa oleosin protein (low molecular weight or L-isoform). Putative regulatory elements are identified and indicated. These include inverted repeats (base pairs 805 to 813 and 821

to 829; base pairs 1858 to 1866 and 1877 to 1885), direct repeats (base pairs 184 to 193 and 1102 and 1111; base pairs 393 to 402 and 1701 to 1710; base pairs 683 to 692 and 1546 to 1555; base pairs 770 to 781 and 799 to 810; base pairs 955 to 964 and 1936 to 1945; base pairs 1483 to 1496 and 1513 to 1526), the abscisic acid responsive element (ABRE) (base pairs 1859 to 1866), CACA box (base pairs 1933 to 1936), TATA box (base pairs 1925 to 1931) and CAT box (base pairs 1989 to 1993). As well, the poly adenylation signal is indicated (base pairs 3020 to 3025). The open reading frame is interrupted by 1 short intron (which are marked) and the 2 exons are translated and indicated in IUPAC single letter amino-acid codes.

Figure 2 and SEQ.ID.NO.:4 shows the DNA sequence of a flax genomic clone encoding a 18.6 kDa oleosin protein (high molecular weight or H-isoform). Putative regulatory elements are identified and indicated. These include direct repeats (base pairs 14 to 25 and 1427 to 1438; base pairs 80 to 89 and 1242 to 1251; base pairs 177 to 186 and 837 to 846; base pairs 1281 to 1290 and 1242 to 1251; base pairs 1591 to 1600 and 1678 to 1287). The open reading frame is not interrupted by introns and is translated and indicated in IUPAC single letter amino-acid codes.

Figure 3 and SEQ.ID.NO.:6 shows the DNA sequence of the flax genomic clone encoding a 2S storage protein. Nucleotide sequencing of this clone revealed it to have an open reading frame of 174 amino acids that showed homology to the plant 2S storage group of proteins. The sequence encodes an open reading frame with 38% overall similarity to a *Brassica oleracea* 2S storage protein, including complete conservation of the glutamine-rich stretch QQQGQQQGQQQ (SEQ.ID.NO.:13). In addition, the 2S storage protein gene promoter contained several putative promoter regulatory elements. These include AT rich repeats (base pairs 25-36, 97-108 and 167-190), RY-like repeat (base pairs 240-247), G-box-like element (base pairs 274-280), seed specific box-like motif (base pairs 285-290) and TATA box (base pairs 327-333).

Figure 4 and SEQ.ID.NO.:8 shows the DNA sequence of a flax genomic clone encoding a 54.4 kDa flax legumin-like seed storage

protein. This legumin-like seed storage protein gene will also be referred to as "linin". The deduced amino acid sequence of the linin gene was compared to the *legumin-like* protein from *R. communis*, the legumin precursor from *M. salicifolia*, *Q. robur* and *G. hirsutum*, the glutelin precursor from *O. sativa* and a 12 S seed storage protein from *A. thaliana*. The linin gene shows a sequence identity/similarity with the corresponding proteins from *R. communis*, *M. salicifolia*, *Q. robur*, *G. hirsutum*, *O. sativa* and *A. thaliana* of 59/15, 47/16, 50/17, 45/17, 43/18 and 43/18 percent respectively. Putative regulatory elements in the promoter region are identified and indicated. These include inverted repeats (base pairs 265 to 276 and 281 to 292; base pairs 513 to 524 and 535 to 545), repeats (base pairs 1349 to 1360 and 1367 to 1378; base pairs 1513 to 1529 and 1554 to 1572), the abscisic acid responsive element (ABRE) (base pairs 1223 and 1231), legumin box (RY repeats) (between base pairs 1223 and 1231), a possible vicilin box region (base pairs 1887 to 1894), CAAT box (base pairs 1782 to 1785) and TATA box (base pairs 1966 to 1970). As well, the signal peptide for ER membrane targeting is indicated (base pairs 2034-2080). The open reading frame is interrupted by 3 short introns (which are marked) and the 4 exons are translated and indicated in IUPAC single letter amino-acid codes.

Figure 5 shows Southern blot analysis of flax genomic DNA. 60 μ g of flax genomic DNA was isolated from leaves, digested with EcoRI (lane 1), HindIII (lane 2) and BamHI (lane 3) and was loaded into the respective lanes. A) Hybridizations were performed with random primed 32 P-labelled 3T cDNA (high molecular weight flax oleosin isoform). B) Hybridizations were performed with random primed 32 P-labelled 7R cDNA (low molecular weight flax oleosin isoform). The results demonstrate that both 3T (high molecular weight oleosin isoform) and 7R (low molecular weight oleosin isoform) oleosin cDNAs hybridize with flax genomic DNA. More specifically the results indicate that 3T is likely to represent a 2-copy gene in flax, as seen by two bands in each lane of digestion. Similarly, 7R is likely to represent a multigene family in flax as multiple bands were detected for each digestion.

EXAMPLE 2

Seed specific expression of flax oleosin genes

Figure 6 shows a Northern blot analysis of the seed specific expression of flax oleosins. Northern hybridization of the two oleosin mRNA in different tissues. Ten μg of total RNA was extracted from different tissues, R, root; C, cotyledon; L, leaf; S, seed capsule; E, embryo. The membrane was probed with (A) cDNA encoding high molecular weight (H)-isoform (identical to coding sequence as presented in Figure 2) and (B) cDNA encoding low molecular weight (L) -isoform (identical to coding sequence as presented in Figure 1). Both the transcripts are expressed only in the embryo and seed capsule, which contains embryos.

EXAMPLE 3

Developmental expression of flax oleosin genes during seed development

Figure 7 shows a Northern blot analysis of the developmental expression of flax oleosins during seed development. 15 μg per lane of total RNA was loaded in each lane on agarose/formaldehyde gel and blotted onto HybondN+ membrane. 10J: This membrane was probed using the ^{32}P dCTP labeled flax oleosin cDNA clone (low molecular weight isoform). Stages indicated are the number of days past anthesis (DPA). 3T) 15 μg per lane of total RNA was loaded in each lane on agarose/formaldehyde gel and blotted onto HybondN+ membrane. 3T: This membrane was probed using the ^{32}P dCTP labeled flax oleosin cDNA clone (high molecular weight isoform). Both the transcripts were expressed very early in development (6DPA, early cotyledonary stage). Expression is maximum at 16 to 20 DPA (late cotyledonary stage) and declines at 22 DPA (mature embryos).

EXAMPLE 4

Transient Seed specific expression of β -glucuronidase (GUS) when under the regulatory control of flax oleosin regulatory sequences

Two constructs were made using standard molecular biology techniques (eg see Sambrook *et al.* (1990), Molecular Cloning, 2nd ed. Cold Spring Harbor Press, including restriction enzyme digestions, ligation and polymerase chain reaction (PCR).

Construct pSC54: The β -glucuronidase reporter coding sequence from vector GUSN358>S (Clontech Laboratories) was placed between the promoter sequence from nucleotide 21 to 1852 and terminator sequence from 2395 to 3501 (as described in Figure 1). This insert was

5 cloned into pBluescript and the resulting vector is called pSC54

Construct pSC60: The β -glucuronidase reporter coding sequence from vector GUSN358>S (Clontech Laboratories) was placed between the promoter sequence from nucleotide 1 to 2023 and terminator sequence from 2867 to 3925 (as described in Figure 2). This insert was

10 cloned into pBluescript and the resulting vector is called pSC60.

pSC54, pSC60 and a promoter-less GUS construct (Control) were introduced into the flax embryos using particle bombardment using standard protocols (eg see Abenes *et al.* (1997) Plant Cell reports 17:1-7). Figure 8 shows the GUS activity of flax embryos bombarded

15 with pSC54, pSC60 and a promoterless GUS construct measured 48 hours after particle bombardment. As can be seen the flax oleosin regulatory sequences are sufficient to drive the expression of GUS in flax embryos.

EXAMPLE 5

Stable seed specific expression of β -glucuronidase (GUS) in flax and Arabidopsis when under the regulatory control of flax 2S storage protein gene promoter

A GUS reporter gene construct was made by incorporating 5' and 3' regions from the DNA fragment described in Figure 3 into

25 promoterless-GUS pBI101 vector as follows.

A 400bp amplicon from the 5' end of the DNA fragment described in Figure 3 was PCR amplified using the following primers (location shown in Fig 3):

5' primer(1): 5'-TCCACTATGTAGGTCATA-3' (SEQ.ID.NO.:14)

30 3' primer(1): 5'-CTTTAAGGTGTGAGAGTC-3' (SEQ.ID.NO.:15)

The PCR primers also contained restriction sites for HindIII and BamHI which were used to clone the 400bp 5'UTR amplicon into the HindIII/BamHI sites of the pBI101 vector in front of the GUS reporter gene. A 736bp amplicon from the 3' untranslated region (3'UTR) of the

DNA fragment described in Figure 3 was PCR amplified using the following primers (location shown in Fig 3):

5' primer (2): 5'-AGGGGTGATCGATTA-3' (SEQ.ID.NO.:16)

3' primer (2): 5'-GATAGAACCCACACGAGC-3'

(SEQ.ID.NO.:17)

The PCR primers also contained restriction sites for SacI and EcoRI. The NOS terminator region of the pBI101 vector was cut out with SacI/EcoRI digestion and replaced with the similarly digested 736bp 3'UTR amplicon of the DNA fragment described in Figure 3.

The GUS reporter construct was then electroporated into *Agrobacterium tumefaciens* strain AGL1 and transformation of flax (Finnegan *et al.* (1993) Plant Mol Biol. 22(4): 625-633) and *Arabidopsis* (Valvekens *et al.* Proc. Natl. Acad. Sci. 85: 5536-5540) carried out according to previously described protocols.

Various tissues from flax and *Arabidopsis* plants carrying the GUS reporter construct were assayed histologically for evidence of GUS activity. In the case of flax, leaf tissue, root tissue and mid-maturity embryos dissected out of developing seeds were stained for GUS activity. For *Arabidopsis*, developing seeds were stained for GUS *in situ* in their siliques.

GUS staining was carried out by immersing the tissues in histochemical buffer containing 0.5 mM X-gluc, 0.5 M potassium phosphate buffer (pH 7.0), 1 mM EDTA, 0.5 M sorbital, 0.5 mM potassium ferricyanide and 0.5 mM potassium ferrocyanide. The staining reaction was carried out for 12-16 hrs at 37°C and the reaction was stopped by adding 95% ethanol. Tissues were subsequently cleared of chlorophyll by repeated washing in 95% ethanol prior to photography. Figure 9 shows clear evidence of strong GUS activity in developing flax embryos and *Arabidopsis* seeds, and no evidence of GUS reporter gene expression in flax roots or leaves, or in *Arabidopsis* silique walls.

EXAMPLE 6

Stable seed specific expression of β -glucuronidase (GUS) in flax, *Arabidopsis* and *Brassica napus* when under the regulatory control of flax legumin-like storage protein gene regulatory sequences

A construct was made using standard molecular biology techniques, including restriction enzyme digestions, ligation and polymerase chain reaction (PCR). In order to obtain a DNA fragment containing approximately 2 kilobases from the 5' transcriptional initiation region of the flax legumin-like seed storage protein in a configuration suitable for ligation to a GUS coding sequence, a PCR based approach was used. This involved the use of the polymerase chain reaction to amplify the precise sequence desired for the expression analysis. To perform the necessary PCR amplification, two oligonucleotide primers were synthesized (Beckman Oligo 1000M DNA synthesizer) have the following sequences:

5' primer: 5'TATCTAGACTCAAGCATACGGACAAGGGT 3' (SJ-634)
(SEQ.ID.NO.:18)

The italicized bases correspond to nucleotide positions 1 to 21 in the sequence reported in Figure 4. The additional nucleotides 5' of this sequence in the primer are not identical to the promoter sequence, but were included in order to place a XbaI site at the 5' end of the amplification product. The XbaI (5'-TCTAGA-3') (SEQ.ID.NO.:19) site is underlined.

A second (3') primer was synthesized which had the following sequence:

3' primer 5'*GGTTATCATTGTATGAACTGA*3' (SJ-618)
(SEQ.ID.NO.:20)

This primer contains the precise complement (shown in italics) to the sequence reported in Figure 4 from bases 2343 to 2363. This primer was not designed with an additional restriction enzyme site due to the fact that a natural NcoI site (5'-CCATGG-3') (SEQ.ID.NO.:21) straddles the start codon between base pairs 2034 and 2039, thereby allowing for insertion of the storage protein promoter into the appropriate cloning vector.

These two primers were used in a PCR amplification reaction to produce a DNA fragment containing the sequence between nucleotides 1 and 2342 of the flax seed storage protein gene with a XbaI site at the 5' end and a NcoI site 302 base pairs from the 3' end. PCR amplification was performed using the enzyme Pfu (Stratagene)

using conditions recommended by the enzyme manufacturer and a temperature program of 94°C (denaturation) for 1 minute, 55°C (annealing) for 1 minute, and 72°C (elongation) for 3.5 minutes. The template was the legumin seed storage protein genomic clone shown in Figure 4.

The resulting amplification product was subsequently digested with XbaI and NcoI to remove the desired 2 kb promoter region. This promoter fragment was cloned into the XbaI and NcoI sites of a XbaI and NcoI digested plasmid designated pGUS1318 (Plasmid pGUSN358S (Clontech Laboratories) was cut with NcoI and EcoRI and the GUS insert was cloned into pBluescriptKS+ (Stratagene) which was adapted to contain an NcoI site in the multiple cloning site.) The resulting plasmid containing the promoter-GUS fusion was called pPGUS1318. The terminator of the legumin seed storage protein from flax was also amplified from the above mentioned genomic clone. To perform the necessary PCR amplification, oligonucleotide primers were synthesized having the following sequences:

5' primer: 5' GCAAGCTTAATGTGACGGTGAAATAATAACGG 3' (SJ620) (SEQ.ID.NO.:22)

The italicized bases correspond to nucleotide positions 3780 to 3803 in the sequence reported in Figure 4. The additional nucleotides 5' of this sequence in the primer are not identical to the promoter sequence, but were included in order to place a HindIII site at the 5' end of the amplification product. The HindIII site (5'-AAGCTT-3') (SEQ.ID.NO.:23) is underlined.

A second (3') primer was synthesized which had the following sequence:

3' primer 5' TAGGTACCTGGCAGGTAAAGACTCTGCTC 3' (SJ-618) (SEQ.ID.NO.:24)

This primer contains the precise complement (shown in italics) to the sequence reported in Figure 4 from bases 4311 to 4290. The additional nucleotides 5' of this sequence in the primer are not identical to the promoter sequence, but were included in order to place a KpnI site at the 5' end of the amplification product. The KpnI site (5'-GGTACC-3') (SEQ.ID.NO.:25) is underlined.

These two primers were used in a PCR amplification reaction to produce a DNA fragment containing the sequence between nucleotides 3779 and 4311 of the flax seed storage protein gene terminator with a HindIII site at the 5' end and a KpnI site at 3' end. Amplification using PCT was as described above. The above pPGUS1318 vector that contains the amplified promoter was digested with XhoI and treated with Klenow to create a blunt end. The vector was subsequently digested with KpnI and the above amplified terminator sequence was inserted so that it was located 3' of the GUS coding sequence. The resulting vector containing the flax seed storage protein promoter, GUS and the flax seed storage protein terminator is referred to as pPGUST.

The XbaI-KpnI insert of pPGUST which contains the linin promoter-GUS coding sequence-linin terminator sequence was ligated into the XbaI-KpnI sites of pSBS3000 (This vector is a derivative from the *Agrobacterium* binary plasmid pPZP221 (Hajdukiewicz et al., 1994, Plant Molec. Biol. 25: 989-994). In pSBS3000 the plant gentamycin resistance gene of pPZP221 was replaced with parsley ubiquitin promoter-phosphinothricin acetyl transferase gene-parsley ubiquitin termination sequence to confer resistance to the herbicide glufosinate ammonium). The resulting vector is called pSBS2089. In addition the XbaI-KpnI insert of pPGUST which contains the linin promoter-GUS coding sequence-linin terminator sequence was ligated into the XbaI-KpnI sites of the *Agrobacterium* binary plasmid pCGN1559 (MacBride and Summerfield, 1990, Plant Molec. Biol. 14 269-276, confers resistance to the antibiotic kanamycin)). The resulting vector was called pSBS2083. Plasmids pSBS2089 and pSBS2083 were electroporated into *Agrobacterium* strain EHA101. *Agrobacterium* strain EHA101 (pSBS2089) was used to transform flax and *Arabidopsis*, *Agrobacterium* strain EHA101 (pSBS2083) was used to transform *Brassica napus*. Flax transformation was performed essentially as described in Jordan and McHughen (1988) Plant cell reports 7: 281-284, except transgenic shoots were selected on 10 μ M L-phosphinothricine instead of kanamycin. *Arabidopsis* transformation was done essentially as described in "Arabidopsis Protocols; Methods in Molecular Biology" Vol 82. Edited

by Martinez-Zapater JM and Salinas J. ISBN 0-89603-391-0 pg 259-266 (1998) except the putative transgenic plants were selected on agarose plates containing 80 μ M L-phosphinothricine. *Brassica napus* transformation was done essentially as described in Moloney *et al.* (1989). Plant Cell Reports. 8: 238-242.

Figure 10 shows the tissue-specific expression of GUS in transgenic flax plants transformed with a linin-GUS gene construct (pSBS2089). GUS expression was measured in roots (R), stems (S), leaves (L), Buds (B) and embryo (E). Some expression was seen in buds, and maximal expression was achieved in embryo tissues. No detectable expression was seen in any of the untransformed (WT) tissues.

Figure 11 shows the temporal expression of GUS in transgenic flax plants transformed with a linin-GUS gene construct (pSBS2089). As can be seen, maximum expression is achieved in mature (pre-dessicated) flax embryos.

Figure 12 shows the absolute expression of GUS in transgenic *Brassica napus* plants (L1 to L9) transformed with a linin-GUS gene construct (pSBS2083). As can be seen high level expression can be achieved in *Brassica napus* plants. When comparing individual transgenic plants, a typical variation in expression due to position effect can also be seen.

Figure 13 shows expression of GUS in transgenic *Arabidopsis* siliques (transformed with a linin-GUS gene construct (pSBS2089)) during seed development. As can be seen high level expression can also be achieved in *Arabidopsis* seed tissues. Maximum expression is achieved at stage 4 (mature but not fully dessicated) of seed development. No detectable expression is observed in non-seed tissues such as leaves, stems, roots and silique walls (results not shown).

While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various

modifications and equivalent arrangements included within the spirit and scope of the appended claims.

5 All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.